Integrated Water Quality and Aquatic Communities Protocol – Wadeable Streams

Standard Operating Procedure (SOP) #23: Data Transfer, Storage, and Archive

Draft Version 1.0

Revision History Log:

Previous Version	Revision Date	Author	Changes Made	Reason for Change	New Version

This SOP explains the procedures for transferring data to the Network Data Manager. In addition, data certification, storage, archiving, and a timeline for project deliverables are addressed.

Data Transfer

All project deliverables, including but not limited to raw data, processed data, Metadata Interview forms, updated data dictionaries, images with metadata, log books, spatial files, and Certification forms will be transferred to the KLMN Data Manager following the timeline listed in Table 1. It is the responsibility of the Project Lead to ensure all products and associated documentation are delivered to the Data Manager following the timeline.

Pre-season Information

At least 3 weeks prior to the start of the field season, it is the responsibility of the GIS Specialist to work with the Project Lead to identify the sites that will be monitored as part of this protocol (SOP #1: Preparations, Equipment, and Safety). Once the sites have been identified, the GIS Specialist should provide the Data Manager with a GIS shapefile that contains the following fields.

- Unique Site Name
- Park Code where site occurs
- Watershed where center of site occurs
- Public Land Survey System Coordinates where center of site occurs
- Name of 7.5’ USGS map where center of site occurs
- Habitat zone for the site (Matrix, Riparian, Alpine)
- Site type (Index, Survey)

Table 1. Deliverable products, responsible individual, due data, and store location for all products developed while implementing the KLMN Stream Monitoring Protocol.

Deliverable Product	Primary Responsibility	Target Date	Instructions for KLMN
GIS Data	GIS Specialist	3 weeks prior to the start of the first field season, and 3 weeks prior to the start of any other field season if changes are made.	Store in Stream_GIS[3]
Contact Information	Project Lead	2 weeks prior to the start of the field season	KLMN Stream Database Stored in the Stream_Data[3] Folder.
Metadata Interview Form	Project Lead	Prior to beginning the first field season and by Feb 1 of the following year.	Store in Stream_Data[3], Use to create and revise full metadata.
Updated Data Dictionary	Project Lead	Prior to beginning the first field season and by Feb 1 of the following year.	Store in Stream_Data[3], Use to create and revise full metadata.
Full Metadata (Parsed XML)	Data Manager	Prior to beginning the first field season and by March 1st of the following year	Store in Stream_Data[3], Upload the Parsed XML Record to the NPS Data Store [1]
Data Certification Report	Project Lead	Every time a product(s) is submitted	Store in Stream_Document[3]
Processed GPS Data Files	Project Lead	Prior to the field crew members being released from service	Store in Stream_GIS[3]
Training Log Book	Project Lead	Prior to the field crew members being released from service	Store in Stream_Document[3]
Digital Photographs and Metadata	Project Lead	Prior to the field crew members being released from service	Store in Stream_Image[3], Copies of Photographs in KLMN Image Library, Copies of Image Metadata into KLMN Image Database linked to Photographs
Equipment Log	Project Lead	Prior to the field crew members being released from service	Store in Stream_Document[3]
Event Log Book	Project Lead	Prior to the field crew members being released from service	Store in Stream_Document[3]
Field Data Forms	Project Lead	December 1st	Store in AQ Office
Scanned Data Forms	Data Manager	December 15st	Scan Original, Marked-up Field Forms as PDF Files and Store in Stream_Document[3]
Initial Databases	Project Lead	December 1st	Store in Stream_Data[3]
Final Database	Data Manager	3 weeks after receiving macroinvertebrate lab data	Store in Stream_Data[3], Send Copy to Parks
Macroinvertebrate Data	Contractor	Prior to the start of the new field season	Store in Stream_Data[3], Upload to the Project Database
Annual Report	Project Lead	March 1st	Store in Stream_Document[3], Upload to NPS Data Store[1], Send Copy to Parks, Post on the KLMN Internet and Intranet Websites, Enter record into NatureBib[3]
Analyses and Synthesis Report	Principal Investigator	Every 3 years on May 1st	
Other Publications	NPS Staff, Principal Investigator	as completed	

[1] NPS Data Store is a clearinghouse for natural resource data and metadata (http://science.nature.nps.gov/nrdata).
[2] NatureBib is the NPS bibliographic database (http://www.nature.nps.gov/nrbib/index.htm).
[3] The KLMN Stream project folder contains five folders: Stream_Documents, Stream_Data, Stream_Analysis, Stream_GIS, and Stream_Image used to separate and store data and information collected as part of the Stream monitoring.

In addition to the site information, it is the Project Lead's responsibility to provide the Data Manager with contact information for each person working on the project at least 2 weeks prior to the start of the field season. Contact information should include:

- First name
- Last name
- Position (Principal Investigator, Network Contact, Crew Leader, Crew Member, etc.)
- Mailing address
- Work number
- Email address

Training

Each crew member should have the appropriate level of training prior to beginning the field work for this protocol. It is the responsibility of the Project Lead (working with the Data Manager, GIS Specialist, and Crew Lead) to develop and implement all components of training associated with this protocol (SOP #2: Field Crew Training). A training log book should be maintained for each crew member and submitted to the Data Manager following the timeline in Table 1. The log book should record the trainer's name, trainee name, position, date, training, and name of equipment. A training log can be downloaded from the KLMN web site or obtained by contacting the Data Manager. For more information on log books, read the KLMN Log Book Guidelines, located on the KLMN web site.

Certification Form

The Klamath Network will utilize a Certification form submitted by the Project Lead to ensure:

1. The data are complete for the period of time indicated on the form.
2. The data have undergone the quality assurance checks indicated in the Stream Monitoring Protocol.
3. Metadata for all data have been provided (when applicable).
4. Project timelines are being followed and all products from the field season have been submitted.
5. The correct level of sensitivity is associated with the deliverables.

A new Certification form should be submitted each time a product is submitted. If multiple products are submitted at the same time, only one Certification form is needed for those products. Certification forms can be obtained from the KLMN web site or by contacting the KLMN Data Manager. An example of the Certification form is included at the end of this SOP.

Field Forms

For the first few years of the project, we will collect data using field forms (in addition to the tablet PC), to ensure we capture all data in an accurate manner. Hardcopy field forms will be provided to the Data Manager following the timeline in Table 1. It is the responsibility of the Data Manager to scan the datasheets into PDF documents within 1 month of receiving the hardcopies. The Project Lead should organize the field forms in the order in which they will be scanned before they are transferred to the Data Manager. Datasheets should follow this order:

1) One PDF will be created for all datasheets for each site.
2) Datasheets should be separated by park.
3) For each site, datasheets should be in the following order:
    a) Stream Verification Form (two sheets)
    b) Field Chemistry and Channel Constraint Form
    c) Torrent Evidence Form

d) Transect Forms (11 or more sheets)
e) Woody Debris Forms (two sheets)
f) Dominant Tree and Invasives Forms (two sheets)
g) Sample Collection Form
h) Vertebrate Specimen Form
i) Discharge Form
j) Slope Form

The scanned document will be named with the park, site name, and the year the data were collected. For example, the scanned document associated with the site 004 of the 2009 stream plot at Crater Lake NP will be: CRLA_004_2009, where "CRLA" is the four letter code for the park, "004" is the site name, and "2009" is the year the data were collected.

A datasheet log should be maintained to record when a datasheet has been misplaced or when a datasheet was destroyed. Since the datasheets will be numbered, a missing datasheet (even if it was thrown out on purpose) must be accounted for at the end of the year. A datasheet log can be obtained from the KLMN web site or by contacting the Data Manager.

Databases

At the end of the field season, following the timeline in Table 1, the Project Lead should transfer a copy of each project database to the Data Manager. The databases should have gone through all validation and verification process outlined in SOP #19: Quality Assurance Project Plan. Once transferred, the Data Manager will subject the data to one more round of validation and verification checks. The Data Manager will work with the Project Lead to correct any errors.

The initially submitted database will only include data entered into the database while in the field and data collected using the *Manta* Multiprobe. Macroinvertebrate data will take longer because of processing time by a contractor and will be uploaded at a later date,as indicated in Table 1. Once the macroinvertebrate data have been obtained from the contractor, the PI should review the data to make certain they are accurate. The PI should then submit the data, along with a Data Certification form, to the Data Manager where they will be uploaded into the project database.

Once the data have been throughly checked, the data will be uploaded into the Master Stream Database, stored on the KLMN server.

Photos

Images and associated metadata will be transferred to the Data Manager in the format explained in SOP #16: Photo Points and Photo Management, following the timeline in Table 1.

GPS / GIS Files

As described above, the GIS Specialist is responsible for submitting a GIS layer of sites that are expected to be visited in a given field season to the Data Manager, 3 weeks prior to the beginning of the field work. These layers will be stored in the GIS folder of the Stream project folder, as shown in Figure 1.

Very little GPS / GIS data are expected to be developed as part of this project. A shapefile of the center point or the reach (Transect F) for each site to be visited as part of this protocol will be

developed prior to the start of the first field season. The GIS Specialist is responsible for submitting a GIS layer of sites that are expected to be visited to the Data Manager 3 weeks prior to the beginning of the field work. Crews will use this shapefile to go to the location on the ground where they will begin to sample the site. Upon the initial visit to the site, a new coordinate will be taken to ensure the accuracy of the location along the stream. These locations will be updated when advancements in technology will provide a better estimate of the location.

Metadata

Following the timeline outlined in Table 1, the Project Lead should submit a Metadata Interview form for the Project Database after each field season. If metadata have not been completed for this project, the form will need to be completed in full. If metadata have been completed, and no updates are needed, just complete question one on the form. In addition to the Metadata form, an updated data dictionary should be submitted following the timeline in Table 1, if the data collected in this protocol have been altered (e.g., new or deleted parameters measured, additions to picklists, etc.). Prior to the start of the following field season, the Data Manager should update the full metadata (Parsed XML) and post it to the proper locations.

Equipment

The field crew should maintain an equipment log to record any updates or changes to the equipment being used to measure the various parameters described in this protocol. It is the responsibility of the Project Lead to submit the equipment log to the Data Manager following the timeline in Table 1.

Reports

There is the potential for a variety of reports and publications to be developed utilizing data collected as part of this monitoring project, including annual reports, Analysis and Synthesis reports, and scientific papers (SOP # 22: Data Analysis and Reporting).

Annual reports will be the responsibility of the Project Lead and should be submitted in the NPS Natural Resource Data Series format, unless utilizing another series format for publication. Final annual reports should be submitted following the timeline in Table 1.

Analysis and Synthesis reports will be the responsibility of the Project Lead and should be submitted in the Natural Resource Technical Report Series format, unless utilizing another series format for publication. Final reports should be submitted to the KLMN following the timeline listed in Table 1.

Data Storage

Project folders have been created for each monitoring protocol the KLMN plans to implement (Figure 1). Project folders contain five standard folders. These folders use a naming convention that includes the vital sign and one of the following: Documents, GIS, Data, Images, or Analysis. These five folders will contain all the data and information for a project as follows:

a) **Stream_Documents.** This folder contains the reports, budgets, work plans, emails, protocols, contracts, datasheets, and agreements associated with a specific project.
b) **Stream_GIS.** This folder contains shapefiles, coverages, layer files, geodatabases, GPS files, GIS/GPS associated metadata, and spatial imagery associated with a project.

c) **Stream_Data.** This folder contains the KLMN Stream database and .dbf files from the six field databases.
d) **Stream_Images.** This folder contains any photographs related to the project and associated image metadata. In addition, copies of all photographs and metadata will be transferred into the KLMN Image database. Details on the KLMN Image database can be found in the KLMN Data Management Plan.
e) **Stream_Analysis.** This folder will contain derived-data and associated metadata created during analysis.

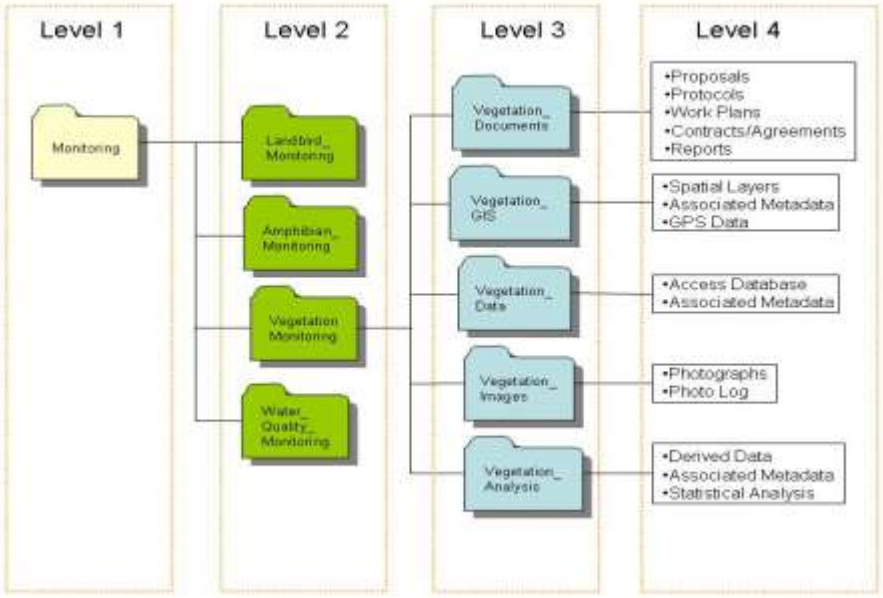

Figure 1. The file structure the KLMN will use to store all stream data and information.

Storage, Backup, and Archiving

The KLMN computer drives are subject to all backup and archiving processes described in the KLMN Data Management Plan. The KLMN relies on Southern Oregon University (SOU) for the backup and long-term storage requirements. Nightly backups are done by SOU to store information that has been edited. This is not a full backup but is intended to protect products that have been manipulated. This information is stored for a 1 week period before it is recycled. SOU begins a weekly full backup of their servers on every Friday and stores the files on tape drives. Backups are stored for 60 days before the tapes are reused. SOU will run quarterly backups on March 31st, June 30th, October 31st, and December 31st of each year. Files stored on a quarterly basis are maintained for 1 year before being recycled (Mohren 2007).

Despite the QA/QC measures in place, finding errors in datasets in the future is inevitable. The process for documenting the correction of such errors is detailed in SOP #19: Quality Assurance Project Plan. In such instances, archived data will not be corrected; however an updated product will be placed into the archive drive along with the digital error and entry logs.

Literature Cited

Mohren, S. R. 2007. Data management plan, Klamath Inventory and Monitoring Network. Natural Resource Report NPS/KLMN/NRR—2007/012. National Park Service, Fort Collins, CO.

Mohren, S. R. 2007. Log book guidelines, Klamath Inventory and Monitoring Network. National Park Service, Ashland, OR.

KLMN Certification Form

1) Certification date: _____

2) Certified by: _____
Title: _____
Affiliation: _____

3) Agreement code: _____
Project title: _____

4) Range of dates for certified data: _____

5) Description of data being certified: _____
_____
_____
_____

6) List the parks covered in the certified data set, and provide any park-specific details about this certification.

Park	Details

7) This certification refers to data in accompanying files. Check all that apply and indicate file names (folder name for images) to the right:

_____ Hardcopy Datasheet(s): _____
_____ PDF Datasheet(s): _____
_____ Database(s): _____
_____ Spreadsheet(s): _____
_____ Spatial data theme(s): _____
_____ GPS file(s): _____
_____ Geodatabase file(s): _____
_____ Photograph(s): _____
_____ Data Logger(s) files: _____
_____ Other (specify): _____
_____ Certified data are already in the master version of a park, KLMN or NPS database.
Please indicate the database system(s): _____

SOP #23: Data Transfer, Storage, and Archive (continued).

8) Is there any sensitive information in the certified data which may put resources at greater risk if released to the public (e.g., spotted owl nest sites, cave locations, rare plant locations)?

_____ No _____ Yes Details:

_____
_____
_____

9) Were all data processing and quality assurance measures outlined in the protocol followed?

Yes / No

If No, Explain _____
_____
_____

10) Who reviewed the products?
_____

11) Results and summary of quality assurance reviews, including details on steps taken to rectify problems encountered during data processing and quality reviews.
_____
_____
_____